

Amendments to the Claims

This listing of claims will replace all prior versions, and listings, of claims in the application:

1. (Currently amended) A method in a data processing system for generating and storing in a database an entry, the method comprising the steps of:

generating an entry comprising:

- i) data identifying a molecule;
- ii) data identifying at least one region in the molecule; and
- iii) a set of axes derived from property distribution information of the at least one region, the set of axes characterizing the at least one region;

generating at least one descriptor vector for the at least one region;

applying a mapping to the at least one descriptor vector associated with the at least one region to construct a key based on preselected association criteria; and

storing the entry in a memory, wherein the key is associated with the entry such that the key indexes the entry for retrieval thereof.

2. (Cancelled)

3. (Cancelled)

4. (Cancelled)

5. (Previously presented) The method of claim 1, wherein the at least one descriptor vector is classified into groups, and wherein the mapping step maps the at least one descriptor vectors to a space discriminating between groups of descriptor vectors.

6. (Previously presented) The method of claim 5, wherein the mapping is derived from the steps of:

generating first data representing differences between groups of descriptor vectors;

generating second data representing variations within groups of descriptor vectors;

identifying a set of component vectors that maximizes a ratio of variations between groups to the variations within groups along the component vectors as a discriminant criterion function;

generating a criterion function for subsets of the component vectors, wherein the criterion function utilizes the first data and the second data;

for each particular subset of component vectors, calculating a probability value for the criterion functions associated with the particular subset;

selecting a probability value from probability values for the subsets of component vectors based upon a predetermined criterion;

identifying the subset of component vectors associated with the selected probability value; and

generating a mapping to a space corresponding to the subset of -component vectors associated with the selected probability value, and storing the mapping for subsequent processing.

7. (Previously presented) The method of claim 6, wherein the first data comprises a matrix \hat{y}_b representing covariance between the groups of descriptor vectors, and the second data comprises a matrix \hat{y}_w representing covariance within the groups of descriptor vectors.

8. (Previously presented) The method of claim 7, wherein the criterion function has the general form:

$$f(\hat{w}) = C \left(\frac{\hat{w}^T \epsilon_b \hat{w}}{\hat{w}^T \epsilon_w \hat{w}} \right)$$

where \hat{w} is some vector, T indicates a transpose, ϵ_b is a first data representing covariance, ϵ_w is a second data representing covariance and C is a constant based upon degrees of freedom in ϵ_b and ϵ_w .

9. (Currently amended) The method of claim 8, wherein ~~the variable~~ C is determined as follows:

$$C = \frac{1/\text{degrees of freedom in } \epsilon_b}{1/\text{degrees of freedom in } \epsilon_w} = \frac{1/(N-1)}{1/(\sum n_i - N)}$$

where represents the number of groups of descriptor vectors, represents the number of regions, and \sum represents the sum of for the N groups.

10. (Previously presented) The method of claim 7, wherein the step of identifying a set of component vectors that maximizes an F-distributed criterion function comprises the substeps of:

determining a set of (eigenvalue, eigenvector) pairs for the matrix \hat{y}_w ; and

determining the set of component vectors based upon the set of (eigenvalue, eigenvector) pairs for the matrix \hat{y}_w .

11. (Previously presented) The method of claim 10, wherein the F-distributed statistic for a given subset of component vectors is based upon value of the criterion function for the subset of component vectors.

12. (Currently amended) The method of claim 11, wherein ~~said statistic~~ the F-distributed statistic for a given subset of component vectors has the following form:

$$\Psi_s = C \left(\frac{1}{L_s} \right) \sum f_k$$

where f_k represents the value of the criterion function at a component vector in the given subset, C is a constant, L_s represents the number of f_k values in the given subset of component vectors, and the Σ operation sums over the L_s f_k values in the given subset of component vectors.

13. (Previously presented) The method of claim 12, wherein the probability value for a particular F-distributed statistic represents a probability value that the particular F-distributed statistic could have been larger by chance.

14. (Previously presented) The method of claim 13, wherein the probability value selected from probability values for the subsets of component vectors is a minimum probability value of the probability values for the subsets of component vectors.

15. (Previously presented) The method of claim 6, wherein the mapping for the at least one descriptor vector performs a loop over each component vector belonging to the subset of component vectors associated with the selected probability;

wherein, in each iteration of the loop, dot product of the descriptor vector with a transpose of a unit vector for the given component vector is added to a running sum.

16. (Cancelled)

17. (Cancelled)

18. (Cancelled)

19. (Cancelled)

20. (Cancelled)

21. (Cancelled)

22. (Cancelled)

23. (Cancelled)

24. (Cancelled)

25. (Cancelled)

26. (Cancelled)

27. (Cancelled)

28. (Cancelled)

29. (Cancelled)

30. (Cancelled)

31. (Previously presented) The method of claim 1, wherein the at least one descriptor vector is invariant to rotation and translation of the at least one region.

32. (Previously presented) The method of claim 31, wherein the set of axes is derived from principal axes of second moments of a region of the property distribution information.

33. (Previously presented) The method of claim 6, wherein the probability value is obtained by treating the ratio as an F-distributed statistic.

34. (Cancelled)

35. (Cancelled)